Context-based Object Recognition: Indoor Versus Outdoor Environments

Ali Alameer[1,2], Patrick Degenaar[1,3], and Kianoush Nazarpour[1,3]

[1] School of Engineering, Newcastle University, Newcastle NE1 7RU, UK,
a.m.a.alameer@newcastle.ac.uk
kianoush.nazarpour@newcastle.ac.uk
[2] School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK
[3] Institute of Neuroscience, Newcastle University, Newcastle NE2 4HH, UK.

Abstract. Object recognition is a challenging problem in high-level vision. Models that perform well for the outdoor domain, perform poorly in the indoor domain and the reverse is also true. This is due to the dramatic discrepancies of the global properties of each environment, for instance, backgrounds and lighting conditions. Here, we show that inferring the environment before or during the recognition process can dramatically enhance the recognition performance. We used a combination of deep and shallow models for object and scene recognition, respectively. Also, we used three novel topologies that can provide a trade-off between classification accuracy and decision sensitivity. We achieved a classification accuracy of 97.91%, outperforming the performance of a single GoogLeNet by 13%. In another experiment, we achieved an accuracy of 95% to categorise indoor and outdoor scenes by inference.

Keywords: Indoor and outdoor classification, Deep learning, Object recognition, Scene recognition

1 Introduction

Over the last decade, machine vision algorithms have reached the level of reliability required in complex environments. In particular, many approaches were advanced for object recognition [1–4]. These approaches share common basis which stands on sifting the input images through a large number of filters to extract features. The extracted features attempt to provide an invariant representation of the object.

Many approaches were developed to achieve invariance to the transformation of objects [5–7]. One successful approach is based on stacking convolutional layers and pooling layers together in a hierarchical structure. Extracting high-level features throughout the advanced layers of this hierarchy has proven successful to achieve invariance. The literature shows many examples of hierarchical structures for object recognition. For instance, the Hierarchical MAX model (HMAX)[8],

consisted of only two stages of convolution/pooling layers. It was inspired by the primate visual cortex. The HMAX model attempts to mimic the processing in the first 100ms of the primates visual cortex, that include processing the visual data through the ventral stream pathway. In this pathway, informative information about the shape and texture of objects that account for rapid categorisation are extracted. Recently, the number of the convolutional/ pooling layers have dramatically increased [9–11]. Increasing the depth of models has increasingly enhanced the classification performance.

Recent studies have shown that models that function well in an indoor environment, perform poorly in an outdoor environment and the reverse is also true [12, 13]. This is due to the stark difference in local and global properties of both environments. The daily life environment, such as living-rooms and city streets, comprises a large number of objects. The nature of these objects depends on the context in which they can be found. Current algorithms of object recognition are trained to recognise objects regardless of their context, dismissing all the information in the backgrounds. This poses a great difficulty for these models to make logical decisions.

Scene understanding is a necessary stage that provides important information about the possible object identity. Identifying the scene can dramatically reduce the probabilities of the object identity and therefore increasing the recognition chance level. For example, outside in a desert, it is more likely to expect a camel than a microscope. We believe that context-based object recognition that depends equally on the environment can characterise the recognition process and therefore enhance the recognition performance.

Hybrid intelligent systems, in particular combining classifiers, can offer a practical solution to handle increasingly complex problems. It allows the use of a priori knowledge to inspire the solution. The concept of hybrid intelligent systems was applied in handwriting recognition, where several neural networks were aligned and a voting process was applied for decision making [14, 15]. It was also shown that averaging the output of an infinite number of independent classifiers can produce an optimal performance [16, 17]. However, the literature has not witnessed utilising hybrid intelligent systems for context-based object recognition, for instance, indoor and outdoor environment.

In this work, we propose three topologies for context-based object recognition. The common factor in all topologies is identifying the environment prior/during the object classification stage. This prior knowledge, i.e., environment type, has given the topologies an advantage in performing classification on a diverse object dataset. We used an object dataset that comprises objects that are likely to be found in an indoor environment. Similar criteria were applied to the outdoor object dataset. We formed three topologies to perform object recognition.

The proposed topologies have the following advantages:

1. It enhances the classification accuracy significantly.
2. It provides more decision confidence. Each decision is based on the posterior probability of more than one classifier (topology-B and topology-C).

3. In topology-C, further to the enhanced classification performance, the object category, i.e., indoor and outdoor, was inferred with an accuracy of 90%.

2 Method

We selected six models of object recognition to form our topologies. We tested them in a challenging diverse visual environment. Below are short descriptions of the models used in this work.

2.1 Shallow models

Here, we refer to the models that consist of five convolutional layers or less as shallow models. Also, we refer to models with more than five convolutional layers as deep models.

HMAX It was inspired by the simple and complex cells hierarchy of the primate visual cortex [2, 8, 18, 19]. It was developed to extract invariant features to object transformations, such as, scaling and translation. It consists of four stages that comprise pooling and convolutional layers. The combinations of convolution and pooling layers are believed to extract a high-level representation of objects.

En-HMAX In the En-HMAX model [20–22], the number of layers was increased. It comprised three convolutional layers and three pooling layers. Additionally, sparse coding and independent component analysis (ICA) were introduced [23, 24]. The ICA method was used to extract Gabor-like filters from natural images in the first simple layer (S_1). Sparse coding was used to train dictionaries in both S_2 and S_3 layers of the En-HMAX model.

AlexNet The AlexNet model [9] is a convolutional neural network (CNN) that consists of five convolutional layers, three pooling layers, and two fully connected layers. It comprises 60 million parameters to fine-tune. It transforms objects in the input images into distinctive features. The AlexNet model operates in a similar fashion to the HMAX model. They share similar hierarchal structure and the same classic alternation of convolutional and pooling layers. Across shallow models, it achieved the highest performances on many datasets [25]. The success of AlexNet has attracted the attention of researchers of computer vision towards CNNs. Due to its simplicity and good performance, in this work, we consider AlexNet, pre-trained with Places dataset [26], as our default model for indoor versus outdoor categorisation task.

2.2 Deep models

Here, we used the following three well-known deep learning CNN models as a major platform to form the topologies. GoogLeNet that comprises 57 convolutional layers is the deepest network used in this work.

VGG16 and VGG 19 The VGGNet architecture introduced in [10], is designed to significantly increase the depth of the existing CNN architectures with 16 or 19 convolutional layers. The last three layers of both versions, i.e., VGG16 and VGG19, are the following layers:

- Fully connected layer: in this layer, the input data is multiplied by the weight matrix and then adds a bias vector;
- softmax layer: in this layer, a softmax function is used for classification purposes. It is considered as the multi-class generalisation of the logistic sigmoid function, also known as the normalised exponential layer;
- classification layer: in this layer, the output predicted label is generated. It is formed by cross-entropy loss function that defines the pre-existed trained classes.

GoogLeNet The GoogLeNet model [11], also known as the inception model, is significantly deeper than the previously explained CNN models. It comprises 57 convolution layers with 5 million parameters to fine-tune. A key feature in the design of GoogLeNet is applying the network in network architecture introduced in [27], in the form of inception modules. Inception module uses a set of parallel convolution layers with a MAX pooling stage along each module. A concatenating layer is used to concatenate the responses of each individual module. In this work, the used version of GoogLeNet comprises a total of 9 inception modules. A more detailed overview of GoogLeNet architecture can be found in [11].

3 Transfer Learning

Transfer learning is increasingly becoming a powerful tool in the field of machine learning [28]. It involves utilising the stored knowledge of a model acquired for solving a particular task and applying it to solve a different problem. For instance, the knowledge acquired while learning to distinguish between different types of trucks could be utilised to recognise different types of cars.

Fine-tuning a network with randomly initialized weights is extremely complicated and time-consuming task. Here, we used networks that were pre-trained with scene images (Places dataset [26]) and object images (ImageNet dataset [25]) depending on the classification task. The CNN models were then adjusted to the new datasets' configurations. To retrain a CNN model on a particular dataset, we froze the weights of earlier layers and only retrained the weights of the advanced layers.

4 Posterior Probability

The posterior probability is the conditional probability that is computed after an occurrence of a relevant event. In the field of pattern recognition [29], the posterior probability indicates the uncertainty of assessing a particular class of

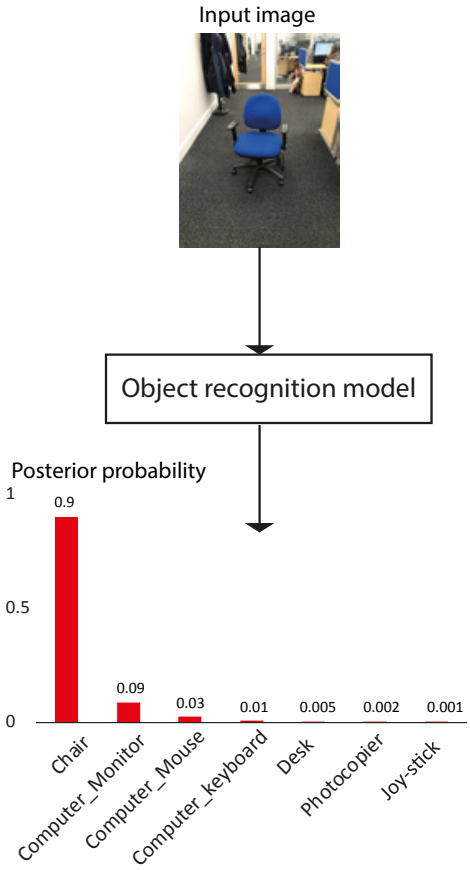

Fig. 1. The distribution of posterior probabilities of an input image. It can be seen that in this example, the classifier is 90% confident that the object in this image is a chair.

images. The posterior probability is produced when a generative model makes a decision [30]. Higher posterior probabilities indicate higher confidence of the classifier's decision. Figure 1 shows an example of how an indoor classifier distributes posterior probabilities for a given input image. Usually, the maximum posterior probability is used to determine the class label. In this work, the maximum posterior probability was utilised to indicate the confidence in the classifier. A threshold for each classifier was set and accordingly, the classifiers made decisions based on their confidence. The threshold was set based on the average posterior probability of all the testing dataset.

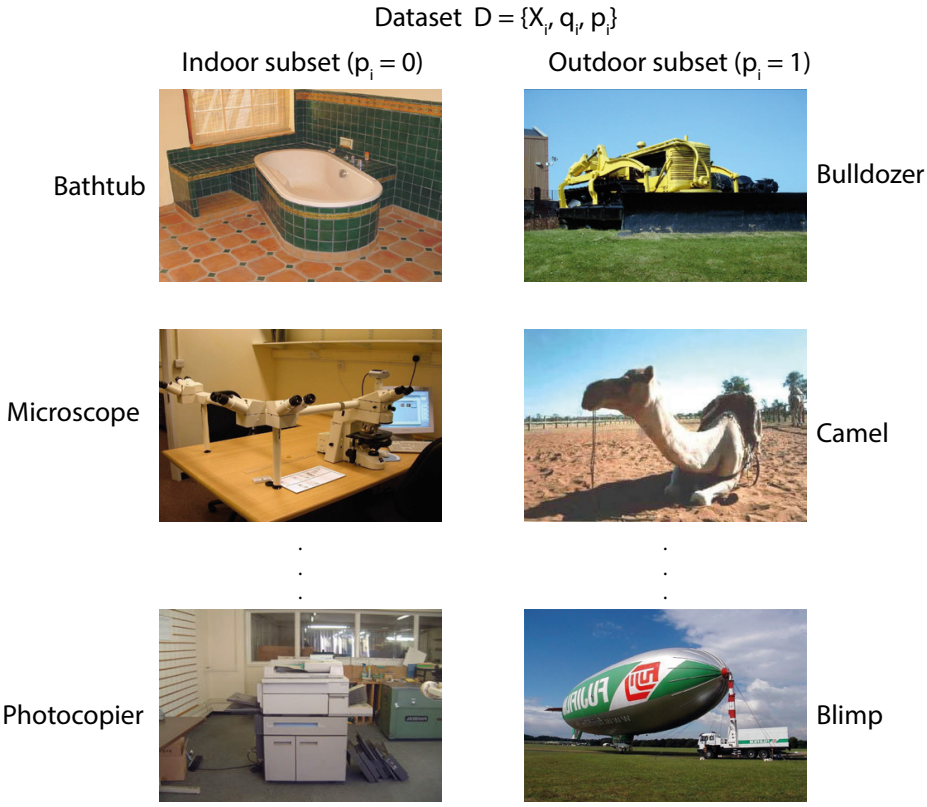

Fig. 2. Selected indoor and outdoor images from our dataset.

5 Datasets

The image classes were collected from ImageNet dataset [25], Caltech 101 dataset [31] and Caltech 256 dataset [32]. These classes were categorised into two uncorrelated set of images: outdoor and indoor. The outdoor image subset does not contain classes of the indoor image subset and the reverse is also true.

Figure 2 shows six examples of the dataset, reflecting the richness of the dataset in terms of the variety of objects and their backgrounds.

6 Classification

In this work, the classification settings are briefly explained. In this section, for all classification scenarios, the extracted features were classified using a linear support vector machine (SVM) [33]. In each of the experiments, 50% of the dataset was allocated for testing the classifier. In addition, to ensure that the classification scores were not biased by the random choice of training samples, the

classification was repeated for 20 runs where the random selection in each round is independent of the other. The average classification score and the standard deviation are reported.

7 Proposed Topologies

The hierarchical topologies developed in this section are designed to achieve an improved classification performance over the existing methods of object recognition. Additionally, providing higher confidence level and decision sensitivity. In this section, a detailed description of the proposed topologies is provided. The method and the architecture of each topology are explained. The designed topologies obtain the environment in which the object is found as an essential component of the recognition process. Furthermore, the designed topologies comprise a decision-making stage that can be tuned to increase the confidence or the decision sensitivity for the process of object recognition.

Topology-A and topology-B consist of three different models for object and scene recognition. They comprise one shallow model for recognising the environment and two deep models for object recognition. Topology-C, however, consists of only two models for object recognition. The environment type, whether indoor or outdoor, in topology-C, is categorised by inference. In this topology, the identity of the environment does not directly contribute to the object recognition process and only computed as an external label.

The architecture of topology-A was inspired by the human visual system, where scenes are rapidly categorised in a small time of 50ms which give a clear information about the identity of the objects within [34]. However, topology-B and topology-C are purely computational with less relevance to biology. Topology-B was designed to minimise the error chance in the first stage of topology-A, the scene recognition stage. The scene recognition stage was designed in-parallel to other stages of object recognition with a different mechanism in the decision-making stage. Topology-C was designed to minimise the number of models in topology-A and topology-B. Only two models for object recognition are used in topology-C for understanding the environment and for identifying objects. Finally, each of the below topologies have several advantages and disadvantages. The below subsections will discuss these in more details.

7.1 Topology-A

Figure 3 shows the basic structure of topology-A. In the used dataset $\mathbb{D} = \{\mathbf{X}_i, q_i, p_i\}_{i=1}^{N}$, each image $\mathbf{X}_i$ has class label q_i (for example: chair) and category label p_i (for example: indoor). The indoor category is denoted by using $p_i = 0$ and the outdoor category by using $p_i = 1$. For a given image, q_i* denotes the predicted class label and p_i* denotes the predicted category label. The confusion matrix of the indoor versus outdoor classifier $C_M = \{c_{ij}\}_{i,j=1}^{2}$ was used to calculate the ratio of the correctly classified images. Using the total probability theorem, the overall accuracy in topology-A can be claculated as shown below:

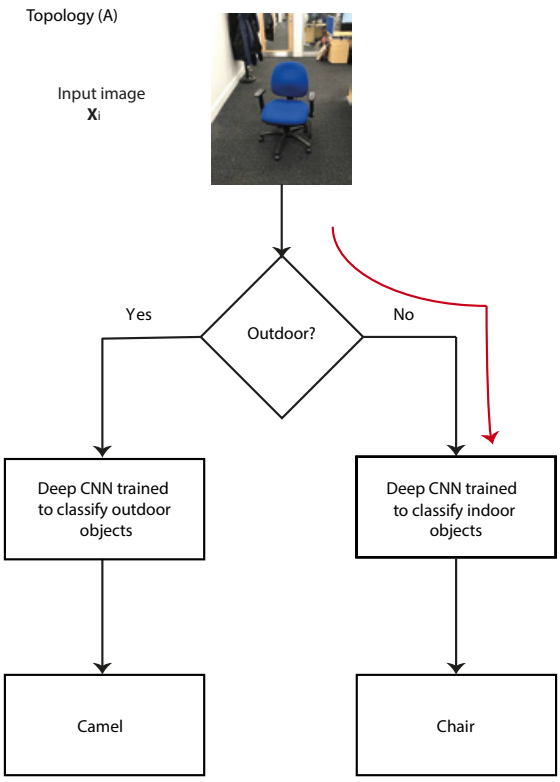

Fig. 3. The structure of topology-A. The input image is first categorised (i.e., indoor and outdoor) then classified (i.e., chair, microscope).

$$
\begin{aligned}
Accuracy(\%) = \frac{100}{\sum_{i,j} \mathrm{c_{ij}}} \; \big[\; & \mathrm{c}_{11} \; \mathbb{P}(q* = q \mid p* = p = 0) \; + \\
& + \mathrm{c}_{22} \; \mathbb{P}(q* = q \mid p* = p = 1) \; + \\
& + \mathrm{c}_{12} \; \mathbb{P}(q* = q \mid p* = 1, p = 0) \; + \\
& + \mathrm{c}_{21} \; \mathbb{P}(q* = q \mid p* = 0, p = 1) \; \big]
\end{aligned}
\tag{1}
$$

7.2 Topology-B

In topology-B, shown in Figure 4, the three classifiers operate in parallel to identify an object in an input image. The object identity depends on the decision of all three classifiers. The three classifiers have an equal influence in making the final decision. Making an incorrect decision in any of the stages does not guarantee an incorrect class label in the final stage. The posterior probability is

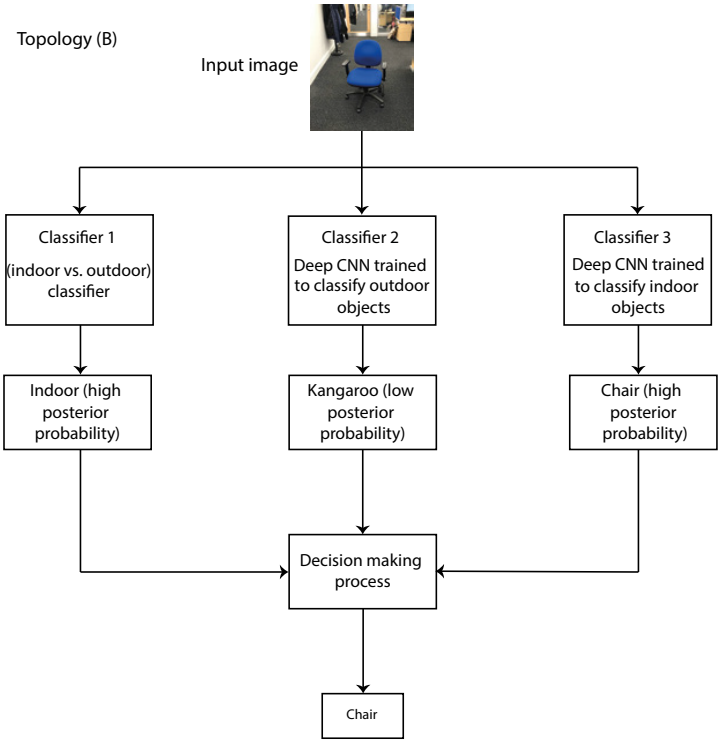

Fig. 4. The structure of topology-B. In topology-B, the classifier that categorises indoor versus outdoor images operates in parallel with other classifiers.

used to quantify the reliability of the classifiers. Classifiers with higher confidence level have more influence on making the final class label decision.

In the experiments performed in this work, the mean of the posterior probabilities of the whole testing data $\mathbb{D}$ was set as a confidence threshold. However, an optimal confidence threshold can be tuned differently depending on the classification context. The final decision is based on the posterior probabilities of all three classifiers as shown in Table 1.

7.3 Topology-C

In this topology, shown in Figure 5, only two classifiers were used to predict the class label and the category label. Table 2 shows the scenarios in which this topology make the final decision.

In this work, the collected image dataset has two separate image subsets. The image classes of the indoor subset do not correlate with the image classes of the outdoor subset. This suggests that when an indoor classifier is used, classes from the outdoor subset tend to give lower posterior probabilities than classes from the indoor subset. Figure 6 shows an analysis of the average posterior prob-

Table 1. The decision-making process of topology-B. The table shows only 2 possible scenarios of the 16th possible combinations. In all other scenarios, a no-decision state will be produced. The ✓marker denotes higher confidence, X marker denotes lower confidence and d denotes the "do not care status".

		Confidence	
Indoor classifier (1)		✓	X
Outdoor classifier (2)		X	✓
Indoor versus outdoor classifier	Indoor decision	✓	d
	Outdoor decision	d	✓
Classifier selection		1	2

Table 2. The decision-making process of topology-c. The ✓marker denotes higher confidence and X marker denotes lower confidence

	Confidence			
Indoor classifier (1)	✓	X	✓	X
Outdoor classifier (2)	X	✓	✓	X
Classifier selection	1	2	No-decision	

ability for both the indoor classifier and the outdoor classifier. In this analysis, GoogLeNet was used to produce the figures. As expected, in both scenarios, i.e., indoor classifier and outdoor classifier, testing a classifier with unseen images within the same training categories produced a significantly higher posterior probability than testing it with different image categories. For the indoor classifier, the Mann-Whitney U test, with a risk $\alpha = 0.05$, shows that the posterior probabilities for indoor test images (M= 87.6, SD = 18.9) were significantly higher than that of outdoor test images (M = 41.7, SD = 21.5); Z-score = 22.3, p-value < 0.05. Similarly, for the outdoor classifier, the above test shows that the posterior probabilities of the outdoor test images (M = 74.0, SD = 26.4) were significantly higher than that of indoor test images (M= 31.2, SD = 18.0); Z-score = 20.9, p-value < 0.05. The data above comprises unpaired non-parametric samples. Therefore, we used Mann-Whitney U method to test for significance. Therefore, we hypothesised that the posterior probability can give a notion of the image category, i.e., indoor versus outdoor.

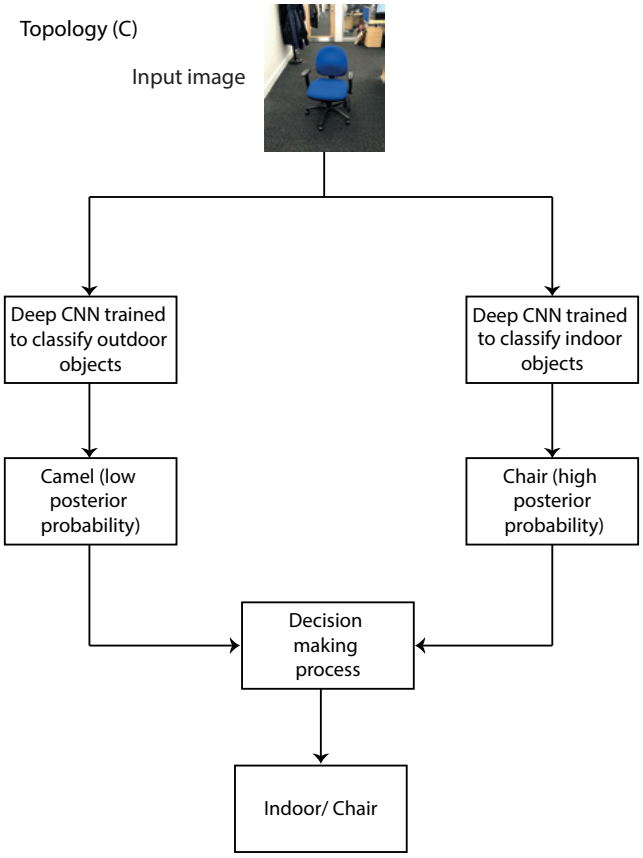

Fig. 5. The structure of topology-C. In topology-C, no classifier is used to categorise the environment (indoor and outdoor), however, it is able to categorise the environment by inference.

8 Results

The below subsections display the results for the discussed topologies in the previous sections.

8.1 Indoor Versus Outdoor

Models of object recognition tend to produce higher performances in a binary classification scheme. The chance level in binary classification scenarios is 50%. In this work, shallow models were utilised for categorising indoor and outdoor scenes. Figure 7 shows a comparison in classification performance between these models. It can be noticed that AlexNet (pre-trained with scene images) outperforms other shallow models for the categorisation task, with a high accuracy of

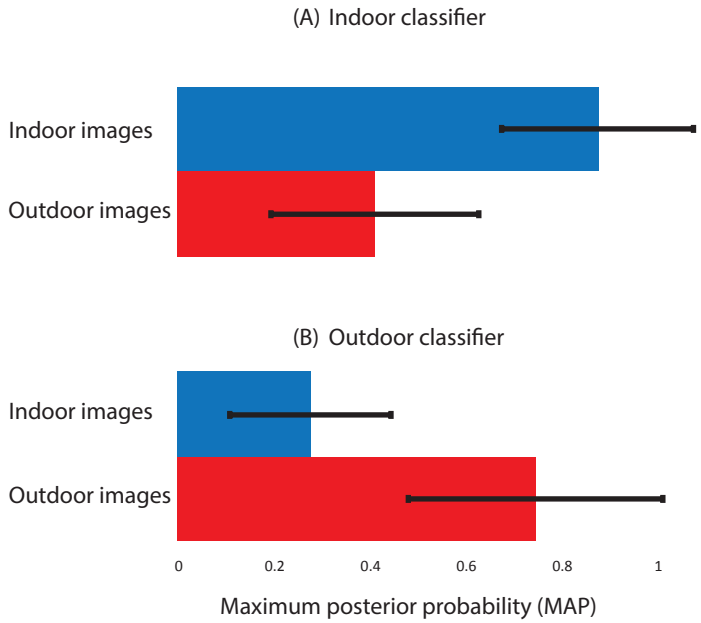

Fig. 6. An example of the average posterior probability of the indoor and the outdoor classifiers using GoogLeNet. (A) Indoor classifier. (B) Outdoor classifier. This chart illustrates the decorrelation in the average posterior probability between the indoor classifier and the outdoor classifier of topology-C.

99.46%. The En-HMAX model achieves higher scores of 87.96%, however, it is still far less than the performance of AlexNet. This is due to the large size of the image data, in which the En-HMAX model cannot handle efficiently due to its abstract architecture. The same applies to the HMAX model, where 75.03% of classification accuracy is achieved. Therefore, AlexNet was elected as a default model with regard to all indoor versus outdoor categorisation schemes, i.e., topology-A and topology-B.

In topology-A, AlexNet spread the images to either the indoor classifier or the outdoor classifier. Although AlexNet has a very high classification performance, the few incorrect decisions it makes lead to failure in the output stage. This is due to the uncorrelated image data used in both classifiers. In another word, the indoor classifier knows nothing about the outdoor environment and the reverse is also true. Therefore, when an outdoor image passes the indoor classifier, an incorrect class label will be guaranteed.

In topology-B, the decision of AlexNet has less impact on the final class label due to the structure of the topology. An incorrect decision at any stage does not guarantee an incorrect class label. In topology-C, however, no shallow network is used to categorise the scene type. The scene type is inferred from the indoor and the outdoor classifier.

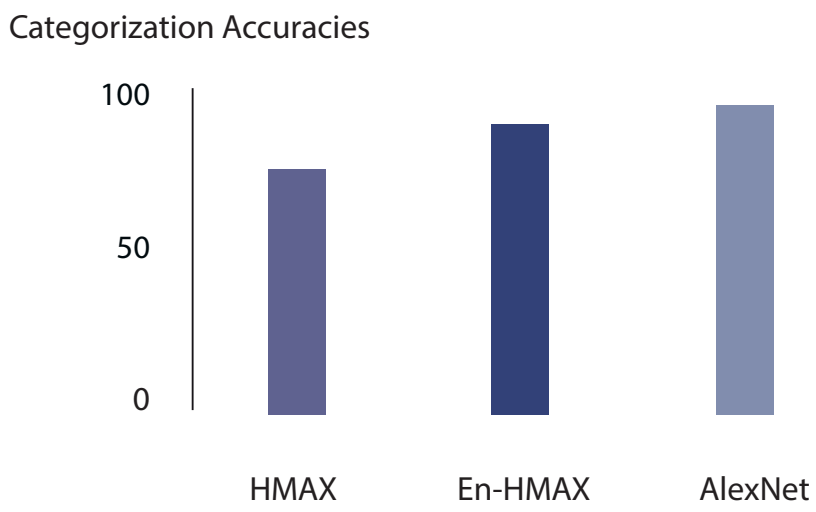

Fig. 7. Results of categorising indoor and outdoor images.

8.2 Classification Scores Using Topology-A

In Figure 8, AlexNet, VGG16, VGG19 and GoogLeNet were utilised as the main platforms to quantify the performance of topology-A. To compute the classification accuracy of the whole classification task, the above models were used individually. In particular, all the image dataset was used without segregating it into an indoor subset and an outdoor subset. This process was repeated for each of the above models separately. As a result, the classification accuracy of each of the above models was quantified for the comparison with topology-A. A similar process was performed for topology-B and topology-C.

Finally, topology-A scores were compared with the above scores. For completeness, the comparisons are only performed between a certain classification model and the topology that is formed within the same model, for instance, the VGG19 network results are compared with topology-A that is formed by only the VGG19 models.

For all used models, topology-A outperformed the original models. For example, in AlexNet, an increased classification performance of 7% is achieved. The difference is constantly decreased for deeper models.

This is particularly interesting because deeper models are capable of understanding large data. Therefore, using a bigger abject dataset is believed to increase the above differences dramatically.

Topology-A has the following advantages:

1. Advanced performance over using a single network.
2. Only two models can operate to recognise each input image.

The disadvantages of topology-A can be summarised as the followings:

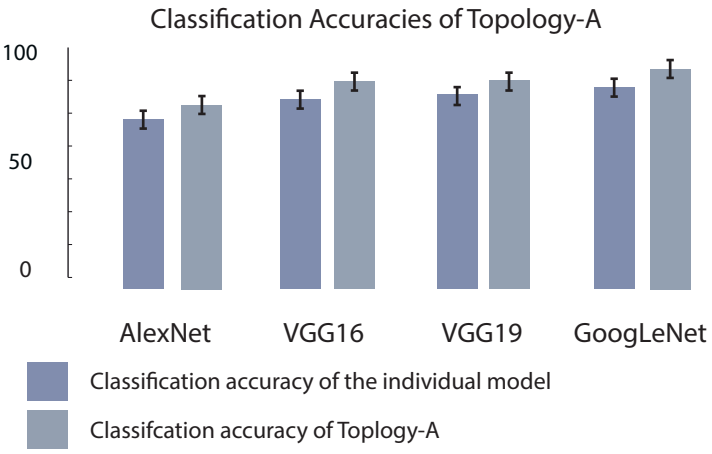

Fig. 8. Results of topology-A. AlexNet is used as a default model for categorising indoor and outdoor images. The classification accuracies in the second-row represent the performance of below models to individually classify the whole dataset.

1. It involves three different classifiers that require more memory in terms of implementation.
2. An incorrect decision in the first stage guarantees an incorrect class label. The first stage (indoor versus outdoor classifier) has more power in making the final decision.

8.3 Classification Scores Using Topology-B

Figure 9 shows the classification scores of using topology-B. It also shows the percentages of the no-decision state. In line with topology-A, similar models were used in this experiment to form this topology. AlexNet was used to categorise the indoor and outdoor images in all scenarios. In the above calculations, the no-decision state is considered as a correct classification. It can be noticed that deeper models such as GoogLeNet and VGG19 do not outperform other models when using this topology. The performances are more balanced. However, the topology formed by VGG19 tends to make more decisions than other models. The decision-making conditions can be tuned using an optimised threshold. In this experiment, the mean posterior probability of all the testing images was used as a threshold of confidence.

Topology-B has the following advantages:

1. The decision-making process depends equally on all three classifiers.
2. It achieves the highest performance among the other topologies.
3. It is designed to make no decisions when a lower confidence level is obtained. The confidence threshold can be tuned depending on the allocated task. Applications with higher risks, for instance, autonomous cars, need higher

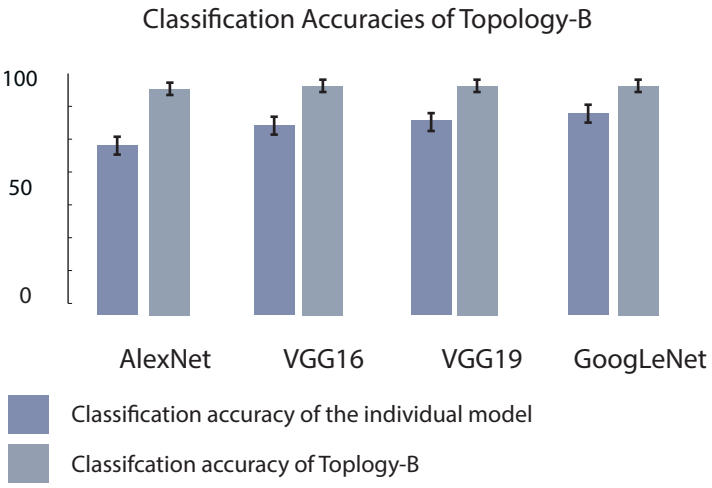

Fig. 9. Results of topology-B. AlexNet is used as a default model for categorising indoor and outdoor images for all the below calculations.

confidence threshold. The "no-decision" state is an important measure in such applications.

The disadvantages of topology-B can be summarised as the followings:

1. It requires more memory in terms of implementation because of the three classifiers in its architecture.
2. It is more computationally expensive than the other topologies because it needs all three classifiers to operate simultaneously.

8.4 Classification Scores Using Topology-C

In topology-C, the objects are classified using only two classifiers as shown in Figure 5. Similar to topology-A and topology-B, the same previously explained models were used to form topology-C. Furthermore, the classification scores were reported in a similar fashion. Unlike topology-B, there was no allocated classifier for categorising the indoor and the outdoor environments. Instead, the category label was inferred throughout the process of recognising an object. Figure 10 shows the categorisation and classification scores of topology-C. A high categorisation accuracy of 95% was achieved using VGG19. This is particularly interesting because this score is achieved without using a specific classifier for the task. In this topology, the percentages of the no-decision state are less than that of topology-B. However, the classification accuracies are slightly decreased. Interestingly, VGG19 performs slightly better than other models using this topology.

Topology-C has the following advantages:

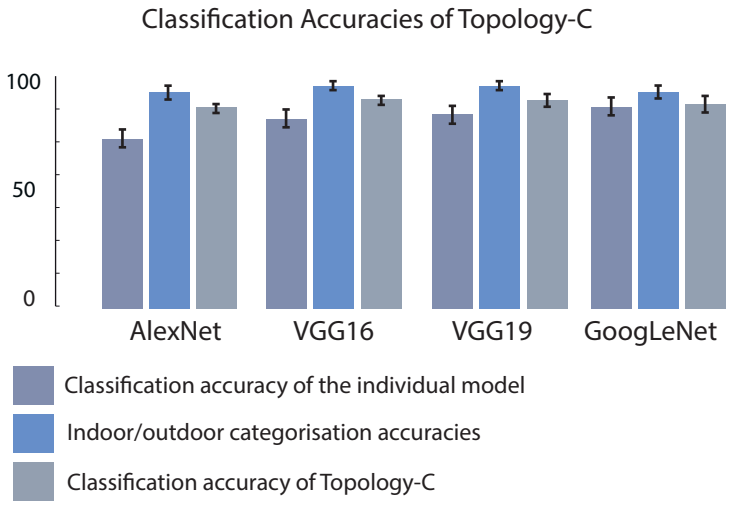

Fig. 10. The results of topology-C

1. It involves only two classifiers for the recognition process.
2. It infers the category label without using a specific classifier, i.e., indoor versus outdoor classifier.
3. It makes no decision when a lower confidence level is obtained.

The disadvantages of topology-C can be summarised as the followings:

1. It provides reduced performance comparing to the other topologies due to the decreased number of the classifiers in its architecture.
2. It shows lower decision frequency than other topologies, due to the limited number of input parameters in the decision-making stage.

9 Conclusions

The architectures presented in this work provide three essential elements for image classification: classification accuracy, decision sensitivity, and computational complexity. In topology-A, two models can operate to recognise objects for each input image. The categorisation stage filters the input images to either the indoor classifier or the outdoor classifier. This topology is less complex than other topologies. However, an incorrect decision at the first stage may guarantee an incorrect image class label. In topology-B, we overcome the problems of topology-A by electing the decision via all classifiers. All three classifiers operate simultaneously and a voting decides the final decision. This topology is computationally complex, as it needs three classifiers to operate simultaneously for each input image. However, it provides higher classification accuracies. Topology-C provides the advantages of topology-A and topology-B. The voting includes only

two classifiers to infer the image category and class. This topology also offers to control the sensitivity of the decision making. Results show that with the proposed topologies, the performance of GoogLeNet can be improved by 13%.

10 Acknowledgements

The work of A. Alameer was supported by the Higher Committee for Education Development, Iraq (HCED, D1201017). The work of K. Nazarpour was supported by the Engineering and Physical Sciences Research Council, U.K., grants EP/M025977/1 and EP/M025594/1.

Bibliography

[1] Z. Li, Y. Wang, J. Yu, Y. Guo, and W. Cao, "Deep learning based radiomics (DLR) and its usage in noninvasive idh1 prediction for low grade glioma," *Scientific Reports*, vol. 7, no. 11, p. 5467, 2017.

[2] X. Hu, J. Zhang, J. Li, and B. Zhang, "Sparsity-regularized hmax for visual recognition," *PloS One*, vol. 9, no. 1, pp. 215–243, 2014.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[4] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of The 22nd ACM International Conference on Multimedia*, 2014, pp. 675–678.

[5] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of Neural Engineering*, vol. 14, no. 3, p. 036025, 2017.

[6] V. Abolghasemi, M. Chen, A. Alameer, S. Ferdowsi, J. Chambers, and K. Nazarpour, "Incoherent dictionary pair learning: Application to a novel open-source database of chinese numbers," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 472–476, 2018.

[7] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "An exploratory study on the use of convolutional neural networks for object grasp classification," in *Proceedings of the* 2^{nd} *IET International Conference on Processing Intelligent Signal Processing (ISP)*, 2015, pp. 5–8.

[8] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, vol. 9, no. 1, 2014.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[12] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 413–420.

[13] A. Alameer, P. Degenaar, and K. Nazarpour, "Biologically-inspired object recognition system for recognizing natural scene categories," in *International Conference for Students on Applied Engineering (ICSAE)*. IEEE, 2016, pp. 129–132.

[14] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.

[15] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[16] K. Tumer and J. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recognition*, vol. 29, no. 2, pp. 341–348, 1996.

[17] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, 1994.

[18] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization," *Proceedings of the National Academy of Sciences*, vol. 104, no. 15, pp. 6424–6429, 2007.

[19] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[20] A. Alameer, G. Ghazaei, P. Degenaar, and K. Nazarpour, "An elastic net-regularized hmax model of visual processing," in *Proceedings of the* 2^{nd} *IET International Conference on Processing Intelligent Signal Processing (ISP)*, 2015, pp. 1–4.

[21] A. Alameer, G. Ghazaei, P. Degenaar, J. A. Chambers, and K. Nazarpour, "Object recognition with an elastic net-regularized hierarchical MAX model of the visual cortex," *IEEE Signal Processing Letters*, vol. 23, no. 8, pp. 1062–1066, 2016.

[22] A. Alameer, P. Degenaar, and K. Nazarpour, "Processing occlusions using elastic-net hierarchical max model of the visual cortex," in *IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. IEEE, 2017, pp. 163–167.

[23] B. Shen, B.-D. Liu, and Q. Wang, "Elastic net regularized dictionary learning for image classification," *Multimedia Tools and Applications*, pp. 1–14, 2014.

[24] A. Hyvärinen, M. Gutmann, and P. O. Hoyer, "Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in V2," *BMC Neuroscience*, vol. 6, no. 1, p. 12, 2005.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248–255.

[26] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in Neural Information Processing Systems*, 2014, pp. 487–495.

[27] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, vol. 6, no. 11, pp. 1019–1025, 2013.

[28] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[29] A. Alameer and H. A. Akkar, "Ecg signal diagnoses using intelligent systems based on fpga," *Engineering and Technology Journal*, vol. 31, no. 7 Part (A) Engineering, pp. 1351–1364, 2013.

[30] F. Ronquist and J. P. Huelsenbeck, "Mrbayes 3: Bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, no. 12, pp. 1572–1574, 2003.

[31] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[32] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[33] V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[34] O. R. Joubert, G. A. Rousselet, D. Fize, and M. Fabre-Thorpe, "Processing scene context: Fast categorization and object interference," *Vision Research*, vol. 47, no. 26, pp. 3286–3297, 2007.